125
165

M U D P I E no.17

Museum and University Data, Program and Information Exchange

THIRD CRAM-COURSE IN TIME-SHARED COMPUTER PROGRAMMING

The third annual cram-course in programming the time-shared computer using the language BASIC will be given on July 5, 1971. The entire day will be devoted to instruction in the language, and July 6 will be used for allowing those who attended the course to use the teletypes in the Museum of Natural History to submit and test programs they have written. The course is aimed at the complete neophyte in computer work; no background or experience of any sort is necessary. Individuals with some knowledge of programming or previous work with time-shared computers are welcome to attend, of course, but it might be a bit boring.

The course will be given in the National Museum of Natural History, Washington, D. C., probably in the divisional library for Reptiles and Amphibians, unless the group gets too large for that room. The library is in Room 207, West Wing. The session will start at 9:00 A.M. It would help if anyone planning to attend would notify J. A. Peters, Division of Reptiles and Amphibians, National Museum of Natural History, Wash., D.C., 20560, as soon as possible, so we can know how many people to expect.--JAP.

NOTE CONCERNING ENCLOSED REPRINT

You will find enclosed with this issue of MUDPIE another paper on a computer topic, written by a Smithsonian staff member, Thomas Waller, of the Division of Invertebrate Paleontology. The compartmentalized mailing lists of the Museum of Natural History are such that papers on computer use cannot be distributed to interested people. The lists are structured to accomodate our divisions, so all people on the invertebrate paleontology list got this paper, whether interested in computer applications or not, but people in other areas and on other lists will not get it unless they request it. Since MUDPIE is sent to people who have either expressed an interest in its purposes, or requested a reprint of a computer paper, it seemed appropriate to suggest that a supply be made available to cover the MUDPIE mailing list. The press agreed, and you have a copy of Waller's paper as a result. If you have no use for it please pass it on to someone who does (and suggest that they write in for MUDPIE, as well), or send it back to me.--JAP.

PETARD HOISTING, No. I

The following excerpt from a letter to a colleague here at the Smithsonian is self-explanatory:

"Enclosed is a copy of MUDPIE no. 15, which describes Dr. Wilsons Monograph on the use of GIPSY in Palynology. If you read the last paragraph, you will note that the author of the newsletter happily pointed out the misspelling of Permian in the Monograph. I think it somewhat ironic (perhaps a touch of poetic justice) that, except where quoted from the Monograph, the word GIPSY is spelled incorrectly. The spelling chosen in the newsletter may be indicative of how the system is used in some cases, however, with all respect to Dr. Wilson, I believe the 'I' spelling would be more closely associated with his use of the system." /s./ Robert W. Shields (U. Okla. Med. Comp. Center).

My old boss, Norman Hartweg, used to say that if you do everything right all the time no one will ever notice, but do something stupid, and it will immediately be brought to your attention. At least I now know that some one out there is reading this stuff!--JAP.

COMMITTEE ON DATA FOR SCIENCE AND TECHNOLOGY

This committee is a subordinate arm of the International Council of Scientific Unions (ICSU), and is called CODATA for short. It is primarily concerned with the compilation of critically selected numerical and other quantitative scientific data, as, for example, "standard heats of formation of water and carbon dioxide, (or) standard entropies at 25 degrees C of selected elements," (from Newsletter no. 1). The feeling of the committee that there was not much data like this in biology is reflected in their first bulletin, which documents "Automated Information Handling in Data Centers," and has no biological centers listed at all. In Newsletter no. 5, however, a paper is included entitled "Critically evaluated data in the biological sciences," by R. L. Zwemer and P. L. Altman, who have worked with the Biological Handbooks Series. This apparently reminded the committee that they had sort of forgotten about biology, and at their meeting in Naples in November, 1970, P. Altman was appointed the representative of the International Union of Biological Sciences on CODATA.

The committee publishes both a newsletter and irregularly issued bulletins. The former deals with miscellaneous information; the latter is usually a vehicle for a single subject, often a report of a "task group" of the committee. CODATA has also published the "International Compendium of Numerical Data Projects," 1969, xxiii + 295 pp., which sells for $13.20. I have not seen this volume, and cannot comment on it.

The newsletter is available on request and free of charge from: CODATA Central Office, Westendstrasse 19, 6 Frankfurt/Main, Federal Republic of Germany.--JAP.

"THE ROCHESTER ULTIMATE WEAPON"

That is what W. Simon calls a new high speed real-time interpretive language developed for use by biologists, and built for operation in a PDP-8 with 4K of memory, requiring DEC tapes. The language is currently available through the Division of Biomathematics, University of Rochester School of Medicine, 260 Crittenden Blvd, Rochester, NY 14620. Although currently only adapted to the PDP-8, it is expected to be functional on the PDP-12 in the near future. The purpose of Dr. Simon in developing the new language was two-fold:

1) To increase the use of small computers by biologists by making it easy for them to learn the fundamentals of computing.
2) To circumvent the problems of machine language use, which was all that had been available on the PDP series previously.

The technique used in the new language is explained in a short note in "Medical and Biological Engineering," vol. 8, 1970, pp. 203-205. A bench mark program took 22 seconds to run in FOCAL, 7 secs in DEC FORTRAN, 6 secs in SNAP (another Simon language), and only 2 secs in the "Rochester Ultimate Weapon." --JAP.

COMPUTERS IN BIOLOGY AND MEDICINE

Volume 1, number 1, of a new journal with the title above appeared in August of 1970. As usual, when people talk about "biology and medicine," they are thinking of biology <u>in</u> medicine. There clearly isn't going to be too much here to interest the ecologist or systematist, although the editor writes, on p. 1, that "the purpose of this journal is to establish an international forum for the exchange of knowledge in the rapidly developing field of computer use in medicine and the biosciences." In a long list of possible subject matter to be covered, including such things as Application of Quantum Chemistry to Molecular Configurations or Functional-force Analysis Applied to Dental Prostheses and other Dental Problems (or maybe you prefer "Computer Aids to Morality," whatever that means), we also find the following MUDPIE possibilities: Taxonomy and Classification Methods; Information Exchange among Research Workers; Applications of Computers to Data Processing in the Biomedical Sciences; Special Purpose Computers for Data Processing; and Computer Programming of Pattern-recognition Analysis. I miss any reference to the possibilities of inter-institution time-shared networks, establishment of common data banks or mutual access storage, and so on. A couple of papers from this first issue are listed below in the "Literature" section.--JAP.

DATA-COLLECTING IN THE SMITHSONIAN

C. A. Bull and R. Shank have completed a survey of the Smithsonian's activities in data collecting, and have summarized their results in an unpublished report entitled "Non-conventional File Structure Data-collecting Projects in the Smithsonian Institution: A Survey, Winter 1968 - Spring 1969." The "non-conventional" designation is for techniques other than simple alphabetical catalogues or lists which will facilitate rapid handling of data, from key-sort to computer. Interest was in projects furthering research or educational functions as opposed to housekeeping. A total of 49 projects were identified, and each is described. Some of the findings are:

Twenty-three projects are using or anticipate using some sort of machine assistance in storage and retrieval.

More projects in biology are using automatic means to store and retrieve data than are projects in the physical sciences, history and technology.

In every case, the highest level of satisfaction with new systems and methodology can be found where the curator and researcher was heavily involved in system design and output control.

The most critical factor is the validity of the data entered and the positive relationship of the degree of expertise of the person making the entry with the reliability of the entry.

In assessing the requirements for making their systems viable, most people who were conducting projects felt the need for manpower with subject knowledge more than the need for electronic muscle.

If anyone would like to know more about the survey, a few copies of the report are available for distribution. Write to MUDPIE for a copy.--JAP.

INFORMATION SYSTEMS SURVEY
"RECON"
(NASA Information System)

The NASA Information System, commonly known as RECON, is a package of programs for creating, maintaining and querying data files. It consists of two major subsystems. The first is a batch system which provides for file creation and maintenance, batch query and prepares output for publication. The second, RECON, is a communication control program, language analyzer and search program which permits multi-programmed query from a remote console.

The system operates on variable length data records composed of fixed header information followed by a variable number of tagged, variable length fields which may be repeated

(i.e., the same field type may occur more than once in a record). Data files are described to the system via Data Description Tables and hence the system is data independent.

The system provides for maintenance and use of a thesaurus for query expansion and for data validation and also permits use of tables for data element conversion to code during file maintenance.

The system was written under contract for NASA. It is operational but is being extended and improved. It operates on the IBM 360/50 under the MFT operating system. The batch program was written in assembly language. RECON was written in PL/I except for the master I/O control program which was also written in assembly language.

FILE ORGANIZATION -- The primary file is made up of data records stored sequentially in accession number order. Each record is composed of a header containing fixed length fields of information which are standard for all records in the file. The header is followed by a series of tagged, variable-length fields with the possibility of repeating a particular tag within the record as many times as is necessary. There is, however, no hierarchy in the sense that a set of fields may be related to each other and repeated as a set relative to the header. The record size is unlimited. The file is stored on a data cell at NASA. It is stored using a specially codecal variable length ISAM access method.

When a file is first defined to the system through a data description table, any field, whether in the header or in the tagged part of the record may be designated for creation of an inverted index. These inverted indexes are stored on disc.

The system also maintains a thesaurus of legal descriptors used for validation during file maintenance. This file also contains "see" and "see also" references and narrower and broader terms. It can be used to expand a query and is available to the requestor at a terminal for browsing.

FILE MAINTENANCE -- In the file maintenance operation, fields in existing records can be altered, records can be deleted from the file and new records can be added to the data base. These are added at the end of the master file. Inverted indexes are maintained by the system as changes occur. Data validation through thesauri and encoding of data via table look-up is also supported.

RETRIEVAL -- The system operates in either of two modes for retrieval: batch or on-line. In the batch mode, records may be selected based upon a Boolean combination of any fields in the record regardless of whether they have been designated as inverted indexes. In this mode also, the query output can be sorted before being presented to the requestor.

In the on-line mode only inverted index fields may be used as terms in the selection logic in order that only qualifying records be accessed from the data cell. No sorting of data is permitted.

Data records may be formatted for output on the XX 1403 printer equipped with an upper/lower case chain, for the Linotron at GPO and for the Photon 713. -- Harriet R. Meadow.

AVAILABLE PROGRAMS

CHECKS--A demonstration program for use with time-shared computers, in BASIC. Written and available from Larry Morse, Biological Laboratories, Harvard Univ., Cambridge MA 02138. Program lists TOTAL CREDIT, CHECKS OUT, and ACTUAL BALANCE.

RECENT LITERATURE

Abrams, M. E. Medical Computing. Elsevier, 52 Vanderbilt Ave., New York NY 10017, 1970: xi + 396. (A series of short papers presented at the conference on medical computing held in Birmingham, Jan., 1969. Re-inforces the notion that medicine and systematics have a lot in common when it comes to computer use.--JAP)

Crovello, T. J. Analysis of character variation in ecology and systematics. Annual Review of Ecology and Systematics, vol. 1, 1970: 55-99.

Goodall, D. W. Statistical plant ecology. Annual Review of Ecology and Systematics, vol. 1, 1970: 99-125.

Ohnacker, G. & W. Kalbfleisch. CCBF -- A System for the Computer Processing of Chemical and Biological Facts. Angewandte Chemie-International Edition in English, vol. 9, 1970: pp. 605-610. (Primarily a chemical system for storage of structural formulae. Biological data is that associated with results of dosage tests.--JAP)

Solow, B. Computers in cephalometric research. Computers in Biology and Medicine, vol. 1, 1970: 41-51.

Walters, R. F., K. P. Brin, F. Roth, J. T. Morrison & C. Renoud. Information support systems for experimental investigation. Computers in Biology and Medicine, vol. 1, 1970: 75-86.

June, 1971
Division of Reptiles and Amphibians
National Museum of Natural History
Washington DC 20560